

Taxim: A Toolchain for Automated and Configurable Simulation for Embedded Multiprocessor Design

Gorker Alp Malazgirt
Department of Computer
Engineering
Bogazici University
34342 Bebek, Istanbul, Turkey
alp.malazgirt@boun.edu.tr

Deniz Candas
Istanbul Gymnasium
34112 Fatih, Istanbul, Turkey
dnzcandas@gmail.com

Arda Yurdakul
Department of Computer
Engineering
Bogazici University
34342 Bebek, Istanbul, Turkey
yurdakul@boun.edu.tr

ABSTRACT

Multicore embedded systems have been constantly researched to improve the efficiency by changing certain metrics, such as processor, memory, cache hierarchies and their cache configurations. Using Multi2Sim and McPAT simulators in combination allows the user to design various multiprocessing architectures and estimate performance, power, area and timing metrics. However, the design time required to simulate these systems is daunting and prone to human error. In this paper, we introduce Taxim, a toolchain that can automatically create requested multicore on-chip topologies along with minimizing the simulation time due to repetitive tasks between architectural power, energy and timing simulations. Taxim's decision-tree-based topology synthesis tool creates processor configuration files that can be highly erroneous when generated manually. The toolchain also automates the steps from design entry to output report extraction by running automation scripts, and listing the results. Our experiments show that multiprocessing architectures with 32 cores and irregular cache hierarchies are more than 1k lines of code in Multi2Sim's processor configuration format and Taxim can create such a file in less than 10 milliseconds. The source code is freely available at <https://github.com/bouncaslab/TaXim/>.

CCS Concepts

•Computer systems organization → Multicore architectures; *Embedded hardware*; Superscalar architectures;

Keywords

Multiprocessors, embedded systems, simulation, decision-trees

1. INTRODUCTION

The current technological challenge faced with embedded systems is balancing for varying domains performance, en-

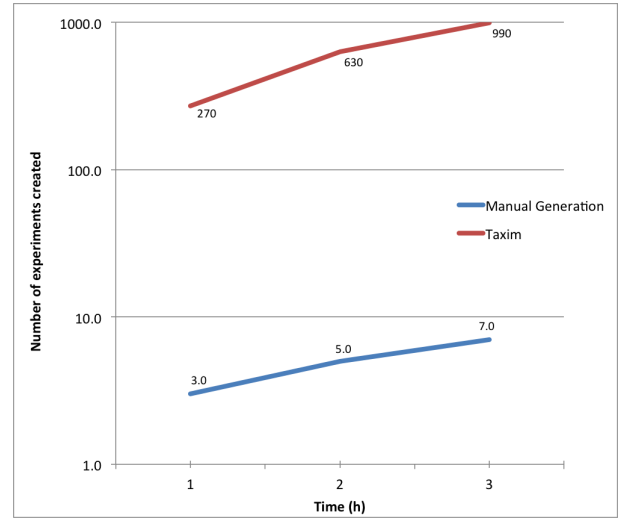


Figure 1: Comparison of manual and automatic experiment generation with respect to time, creating 4-core multiprocessor topologies

ergy efficiency, and production costs that are dominated by the design time and the chip area. The simulation tools are essential for designing complex systems because these tools help the designer in searching the design space and without experimenting on physical implementations, which are time consuming and costly. In the embedded systems design domain, cache [16], memory [15], architectural [18],[3] and energy [12] simulators exist so as to evaluate the energy or performance of processor architectures in the large design exploration space. Thus, simulators are coupled with design space exploration (DSE) tools [22],[7] where optimal designs are sought using algorithmic approaches based on designer's requirements. In general, this procedure consists of a large number of iterations between DSE's optimizer and the simulators, and continues until sufficiently good results are achieved. Hence, in general, at each iteration the design under simulation undergoes hardware or software modifications and this process requires automation.

Multiprocessor design exploration requires generation of various multiprocessor topologies. These topologies do not only differ in cache organizations and number of cores, but they also vary in the number of caches, cache levels, and data/instruction cache sharing metrics. Specifically, for mul-

tiprocessing with many cores and caches, manual cache-processor-memory connectivity creation is time consuming and prone to human error.

In this work we present Taxim, a toolchain that automates architecture generation and the simulation of embedded multiprocessing systems using Multi2Sim [18] and McPat [12] simulators that have been used extensively in research and DSE tools. Multi2Sim offers the designer a fully customizable environment that simulates both the devices and their interactions. McPat on the other hand allows a detailed exploration of crucial efficiency metrics.

Automated architecture generation is fast and not prone to factors such as fatigue. In Figure 1, we present the number of experiment creation with respect to time. Experiment creation task includes the generation of 4-core multiprocessor topology lists that have less than 200 lines of Multi2Sims [22] architecture representation files, automation scripts and simulation output preparation. Based on our observations on a group of three students, we have measured that the rate of design generation decreases over time due to fatigue. Thus, However, with an automated tool, after an initial setup time to specify design metrics and prepare a topology list, experiment preparation continues at a constant pace.

In this paper, Taxim’s contributions can be summarized as:

1. In Taxim, we introduce a decision tree based method for that creates processor multi-core processing architectures by using decision trees. Thus, Taxim speeds up design time considerably when compared with the manual generation of the architectures.
2. It provides an automated toolchain that allows designers to run multiple simulations using Multi2Sim [18] and McPat [12]. In addition, Taxim is built to prepare structured output reports of numerous results such as processor datapath specifics, cache line utilization to leakage power consumption. Design space exploration (DSE) tools [22, 7] can utilize these reports.
3. While setting up the environment for multiple simulations, the designer might enter some parameters wrong. Hence, the simulations might have to get aborted. Yet, the existence of erroneous simulation set-ups might truly slow down the completion time of multiple simulations. In order to prevent erroneous simulations due to human factor, we introduce intelligent control in Taxim. In this way, multiple simulations can be carried out smoothly in the shortest possible time.

We present Taxim in the following way. We first present related works in the next section. Section 3 details our architecture generation and validation method. In Section 4 we detail Taxim’s process flow. Section 5 presents the experimental results that compare Taxim and manual approaches. We conclude the paper in Section 6 with our conclusions.

2. RELATED WORK

Simulators have been an important part of designing processing systems [3, 20, 6, 14, 17, 8]. The first step before running a simulation is design entry, that allow designers to customize the hardware and the benchmark that will execute. Our work differs from available works as follows:

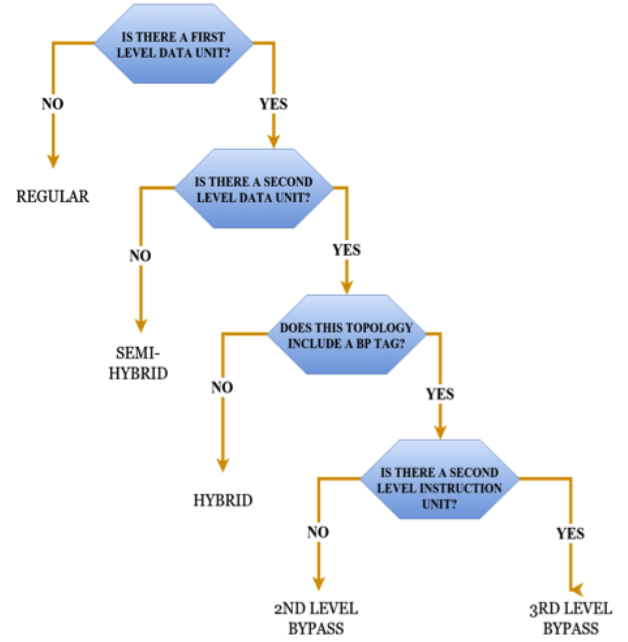


Figure 2: Decision tree that represents different pathways taken by Taxim to create a given topology

First, we present an automated processor architecture generation method based on a decision tree. It allows us to generate very complicated multi core architectures with multiple levels of cache hierarchy with ring bus network support when necessary. Automatic architecture generation tool also validates cardinality and connectivity of generated architectures. This allows automatic generation more advantageous over manual generation specifically for complex designs.

Second, we present an automatic toolchain, which automates the step from design entry to output, report extraction for architectural and power, area and timing simulations. In this work, unlike existing research works we present these steps decoupled from other topics such as design space exploration [21, 19, 10] or simulation methods [19]. In this way, our tool-kit can be used as a plug-in by the design space exploration tools. Since the environment enables multiple concurrent simulations, the simulation time can be considerably reduced in multi-core machines and supercomputers. Obviously this will reduce the design exploration time when Taxim is coupled with DSE tools.

In the design space exploration literature, there have been plenty of studies that require large number of simulations for finding pareto set or the single “best” design point from single or multiple design parameters [6, 17, 19, 10, 1, 11, 21, 5, 8]. For this reason, this work exhibits an automatic architecture topology generation method that can be used in DSE systems that employ Multi2Sim [22] and McPat [4]. Alongside automated architecture generation, automated design entry to extraction of output reports allows DSE tools to process design points in a rapid and continuous manner. The authors in [1] explore energy consumptions of various two level cache configurations using McPat simulations. There has not been any automation regarding the architectural simulations due to the fact that the underlying processor ar-

chitecture is fixed. Similarly, the work in [11] explores cache hierarchy by estimations and simulations. However, cache topology has been fixed, thus automated cache topology generation has been omitted. However, our work first presents automatic architecture generation, then explains the rest of the tasks that automate whole simulation process from design entry to output extraction, using Multi2Sim and McPat environments.

Authors in [6] have presented a design space exploration of computer architectures that searches in a large design space. The automated tool customizes cache architectures and processor internals, however there has not been any architectural exploration such as connectivity of different cache levels, cache sharing and processors in Multi2Sim and McPat. In addition, the generation scripts have only been limited to single-core architectures. The work in [8] takes a range of high and low-level parameters to improve accuracy in the design of a multiprocessor system on a chip. Authors automate generation of processor specifics such as the instruction set, number of pipelines and pipeline stages. However, cache topology is fixed whereas our architecture generation tool also automates the cache topology creation.

Cache hierarchy exploration engine presented by Yessin et al [21] first determines the boundaries of the design space by determining the ranges of the design variables and prepares the memory traces through simulation. Next, an optimization algorithm iteratively determines the most suited cache hierarchy from the given benchmark set. By using an automated toolchain similar to Taxim, the optimization unit of the system can request additional simulations, thus the feedback loop can be expanded with more simulation results. Similarly, the work in [19] provides an mpsoc simulation environment where designers can provide automation scripts to generate simulations with various design parameters via the parameter file. The presented environment allows embedding different memory units with different Network-On-Chip (NOC) architecture support. In this way, the simulation tool can be connected to various design automation tools. The design space can be enlarged by combining simulation and estimation of architectures with tolerable error margins. The authors in [10] also provides a method for performance estimation of pipelined multiprocessor system-on-chip architectures. Analytical models are combined with simulation data and exploration is handled on the aggregated design space. The design of the experiments requires numerous experiments therefore both simulation and estimation requires architectural preparation and application settings. Thus, when the number of architectures and benchmarks increases, a toolchain like Taxim plays an important role for automating architectural preparation, simulation/estimation and results extraction tasks.

3. DETAILS OF ARCHITECTURE GENERATION

In this section we detail the architecture generation and validation methods. Our nomenclature is presented in Table 1. We define the number of cores with C. When data and instruction caches are shared, we represent these caches as L. When instruction and data caches are separate, they are specified as IL and DL respectively. If there exists a bypass connection [13], it is shown with BP. We can show usage of the nomenclature in two examples:

Algorithm 1: Architecture generation from string representation

Input: An architecture's string representation

Output: Multi2Sim processor configuration file

Step 1: Parse given architectural string and tokenize cores, instruction caches and data caches;

Step 2: Determine number of cores, then determine the number of first, second and third level instruction and data caches, group them with respect to levels;

Step 3: Traverse the decision tree and select architecture type from available types: (Regular, Semi-Hybrid, Hybrid, 2nd Level Bypass and 3rd Level Bypass);

Step 4: For each core in the system:

Connect each core to an empty first level data cache;

If there does not exist an empty first level data cache;

Connect each core to the least connected first level data cache;

Connect each core to an empty first level instruction cache;

If there does not exist an empty first level data cache;

Connect each core to the least connected first level instruction cache;

Step 5: For each level in the cache hierarchy:

Connect each data/instruction cache to an empty upper level data cache

If there is no empty cache;

Connect each core to the least connected upper level data cache;

If there is no upper level cache;

Connect each cache to the memory;

Step 6: If BP exists in the string representation:

For each last level memory component connect a network switch:

For each incoming cache connection to each: last level memory

For each cache level in the system do separately for each data and instruction cache:

Group data/instruction cache connections in two and connect to a switch;

Connect each switch pair with a new switch;

If total number of caches is odd, connect the remaining cache to one of the pairing cache switch;

Connect each switch pairs to the memory;

Symbol	Definition
C	Core Count
L<x>	Shared Data/Instruction Level 1, 2 or 3 Cache
IL<x>	Instruction Level 1 or 2 Cache
DL<x>	Data Level 1 or 2 Cache
BP	Exists a bypass connection

Table 1: Nomenclature of the Topology Representation

- 2C_4L1_1L2 explains that there exists two cores, four L1 shared caches and one L2 shared cache
- 2C_2DL1_2IL1_1DL2_BP shows us that there exists two cores with two data L1 cache, two instruction L1 caches, one data L2 cache and there exists a bypass connection from instruction caches to the memory

This nomenclature is followed to create topology representations of designated architectures. The designer can enter in string format to receive configurations recognizable by the simulator. The designer can also opt to work with provided topology sets that are available in the library of the toolchain. The topology creation tool relies on strictly defined "Topology Creation Rules" as outlined in [13]. The user may specify the desired topologies in a text file by entering their names according to the designated nomenclature.

Architecture Generation: During architecture generation, Taxim first tries to extract the type of topology by parsing the names of the architectures. The parser also checks compliance of the architecture name to the "Topology Generation Rules" [13]. Common design rule errors are missing underscores, wrong order of caches, missing or extra letters that do not conform to generation rules. After parsing the nomenclature without any errors, a simple decision tree is employed as shown in Figure 2. The decision tree (DT) guides the generation tool for the type of routing that should be applied to construct the given topology.

The decision tree consists of four rules to call five different methods. These methods are: Regular Fat Tree, Semi-Hybrid, Hybrid, 2nd Level Bypass and 3rd Level Bypass. The generation rules can be summarized as follows and they are derived from "Topology Creation Rules" [13]:

- Rule 1: Taxim checks if there exists separate data and instruction caches. If not, the generator employs Regular method where Von Neumann or Harvard architectures can be employed
- Rule 2: Taxim checks if there exists second level data unit. Absence of second level data caches employs Semi-Hybrid method. In this method, the generated architectures are hybrid architectures because there exists shared instruction caches at the first level. Nevertheless, above first level, all data and instruction caches are shared without any bypassing
- Rule 3: Taxim checks if given topology representation has BP tag that stands for by-passing a cache. Absence of BP tag means that the given architecture is a hybrid architecture and data caches are separate from instruction caches at all levels
- Rule 4: Taxim checks if there is a second level instruction cache. The absence of second level data cache employs the 2nd level bypass method that generates architectures with first level instruction caches bypassing second level cache to main memory. Finally, the 3rd level bypass method is employed that means second level instruction caches are bypassed third level caches.

The method types, brief explanations and a two core sample topologies are shown in Figure 3. These methods are derived from "Topology Creation Rules" [13]. The advantage

of decision tree is to extract the maximum available information from the string representation and separate domain information parsing from architecture generation. In Algorithm 1, we show the architecture generation procedure of Taxim.

Algorithm 1 starts by parsing the input that is the topology string representation. After string representation is parsed, Taxim tokenizes the cores, instruction/data caches and their levels. Based on this information, the decision tree is traversed and type information of the topology is extracted that is shown in Line 3. Then, the algorithm proceeds by connecting each core to first level data caches. Based on available data caches and instruction caches, each core is either connected to an available cache, or to a cache that has the least amount of connection. After all cores are connected, first level cache connections occur. Each cache level connects to an upper level cache that is free or has the least amount of connection. At the highest cache level, each cache is connected to the main memory. These steps are represented between Line 4 and 5. Based on these steps, the connectivity of the topology can be defined as the following:

Let nc be the number of components at level i , mc be the number of components at level y such that $i < y$, and from Topology Generation Rules [13], $nc \geq mc$

Components at level i must be connected to components in level y and after each component at level i is connected to an available component at level y , the number of connections c_{iy} created by Algorithm 1 becomes:

$$c_{iy} = \lfloor nc/mc \rfloor \quad (1)$$

If the number of components is odd, the remaining connections are connected injectively, and the number of remaining connections cr_{iy} become:

$$cr_{iy} = nc \% mc \quad (2)$$

Thus, the total number of connections between level i and y ct_{iy} becomes:

$$ct_{iy} = c_{iy} * mc + cr_{iy} \quad (3)$$

After core/cache connections are established, Taxim checks if there are any bypass representation in the given topology representation in Line 6. If there exists any bypassing, Taxim starts to add network switches for forming the bypass structure. The first step is to connect a network switch to the last level memory such as the main memory or last level caches. Then, it starts grouping data caches into pairs. If the number of data caches is odd, the last remaining cache is connected to a pair making which makes it a triple. Then, a new switch is added to an each pair of caches. Similarly, each switch pair is connected to a new switch until all switches are covered. If total number of cache pair switches is odd, it connects to an existing cache switch pair, making it a triple. After data cache switches are created, same steps are applied for instruction caches as well. When all data/instruction cache switches are generated, these switches are connected to the main memory switch.

Figure 4 visualizes the network creation process explained in Algorithm 1. After Taxim connects DL2 and IL1 caches to main memory in Figure 3A, first a switch is connected to the main memory. Then, DL2 caches are grouped and they are connected to a switch. Next, IL1 caches are also grouped and connected with a switch. The main memory switch is then connected with the cache grouping switches.

Type	Explanation	Example	Figure
Regular	Regular Harvard or Von Neumann Architecture and can have up to three levels of cache hierarchy	2C_2L1_1L2	
Semi-Hybrid	Only at the first cache level, data caches are dedicated with a shared instruction cache. In the upper levels in the memory hierarchy, instruction and data caches are unified.	2C_2DL1_2IL1_1L2	
Hybrid	At all levels data caches are dedicated. Instruction and data caches are not shared.	2C_2DL1_2IL1_1DL2_1IL2	
2nd Level Bypass	There are no second level instruction caches present in this configuration, as IL1 caches are immediately connected to the upper level (either L3 or Mem)	2C_2DL1_2IL1_1DL2_BP	
3rd Level Bypass	Generated only if there are L3 caches. The IL2 units skip the Level 3 and are connected to the main memory unit directly	2C_2DL1_2IL1_1DL2_1IL2_1L3_BP	

Figure 3: Comparison and explanation of topology methods and their examples

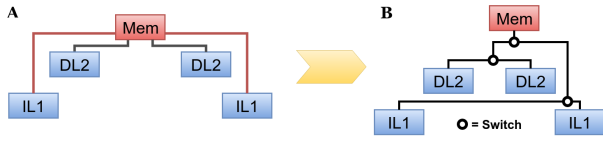


Figure 4: Logical representation (left) and implementation (right) of the network of 5C_5DL1_2IL1_2DL2_BP

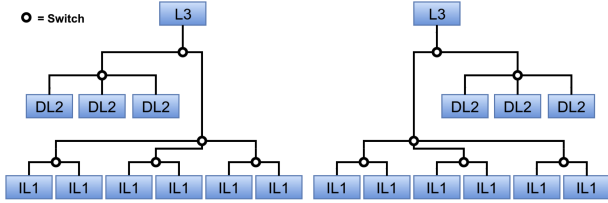


Figure 5: The switch network connections for a bypass structure, there exists no more than three connections from lower level to upper level in the hierarchy

In Figure 5, network switch connections of a more complicated topology, 24C_24DL1_12IL1_6DL2_2L3_BP (a second level bypass) is shown. First, an L3 cache is connected with a switch. Pairs of IL1 caches are connected with a switch and three IL1 switches are connected to a switch and then connected to the L3 switch. Three DL2 switches are connected with a switch and then this switch is connected to the L3 switch. These connections are repeated for the second L3 cache, starting with creating a switch for the L3 cache.

As a design example, let us assume that the designer requests the creation of 5C_5DL1_2IL1_2DL2_BP which is shown in Figure 6. According to the decision tree, first rule is to check whether this topology has a first level data cache (DL1) component. If it is available, the requested topology can be a regular type. Next, the system checks whether a second level data cache (DL2) component also exists and if not, it would be a semi-hybrid memory configuration. In the third rule, it is checked whether the suffix BP has been added to the nomenclature. If true, it is checked whether there is a second level instruction unit. Since in our case there is no second level instruction unit, the system identifies that we have a type "Level 2 Bypass". The bypassing occurs from second level instruction caches to the main memory.

The topology generation algorithm connects each core in the architecture to a single DL1. Then, each core is connected to a free IL1, however there are only two IL1 in the configuration. Hence, Core 1, 3 and 5 are connected to one IL1 and Core 2 and 4 are connected to the other IL1 as designated by the algorithm. After the cores are connected, first level data caches are connected to second level caches. There are two DL2 in the topology, therefore three DL1 are connected the first DL2 and the other two are connected to the second DL1. Then, there are not any third level caches in the topology, therefore all DL2 are connected to the memory. Since there is no IL2 in the topology, all IL1 are connected to the memory, bypassing second level caches. Checking component cardinality and connectivity of the generated architecture validates the generated topologies. After

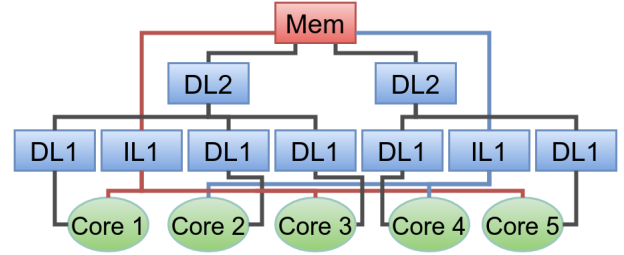


Figure 6: Topology of 5C_5DL1_2IL1_2DL2_BP Architecture

given topology string representation is tokenized and cores, data/instruction caches in every cache level are extracted, cardinality of these tokens are compared with the cardinality of the generated cores and caches. For determining the connectivity of the caches and cores in general, the validation method traverses each component in the generated architecture topology and computes the number connections described in Equations (1), (2) and (3).

As an example, in Figure 6, there exists 5 Cores and 5 DL1 caches, thus each core is connected to a DL1 cache. However, there are 2 IL1 caches, thus after each cache accepts $\lfloor 5/2 = 2 \rfloor$ connections, the remaining $5\%2 = 1$ connection must be connected to the first available cache in the cache list, increasing the number of connections to 3 for one of the IL1 caches. Similarly, at the second cache level, the connectivity between 5 DL1 and 2 DL2 are validated the same way.

4. TAXIM TOOL FLOW

In this section we present the process flow of Taxim and detail the important parts. Figure 7 shows the process flow. The first step is defined as design entry. In this step the designer determines design and simulation specific properties. In the second step, Taxim's stencil scripts read design entry files and create the simulation environment and architectural topologies that are used by Multi2Sim and McPat [12]. Third step is composed of simulation with intelligent control that allows early termination of poor performing design with respect to a user determined design parameter such as latency. The last step includes extraction of simulation results. Results are also stored for later usage for analysis purposes.

The outlined steps are intended to guide the user from designer specifications to aggregated results of numerous parameters that are selected by the designer. As the diagram displays, the designer requirements regulate most of the process flow. The input files are tailored according to chosen topologies and benchmarks, which allow Multi2Sim and McPat [12] to examine their effectiveness. The acquired data is filtered through selected parameters to extract the data for the further evaluation of the designer.

4.1 Design Entry

In this step, designer enters all the files that Taxim needs for running simulations. As shown in Figure 8, these are the benchmarks, their input data sets, processor architectures and design parameters that user aims to collect. Benchmarks should be entered in executable form with their required input data sets. Unlike some other platforms in the

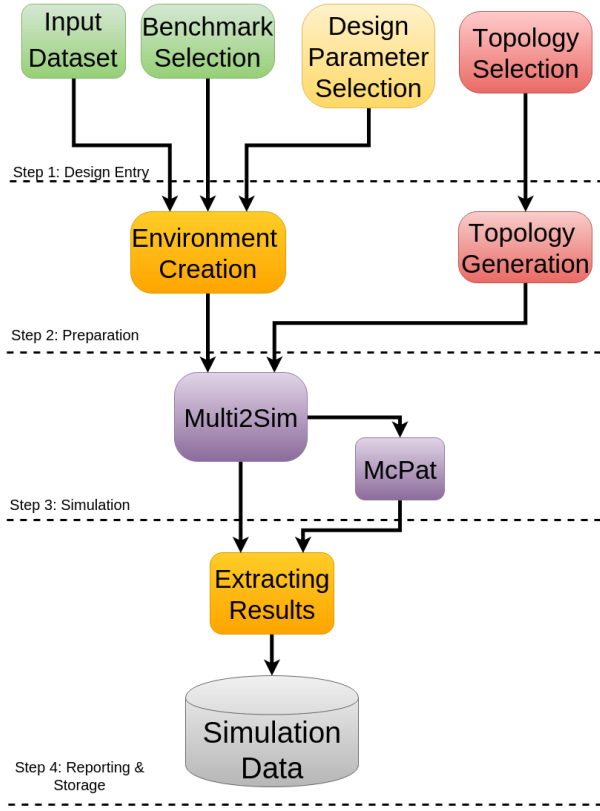


Figure 7: Process flow of Taxim representing all steps from user input until results extraction

Component	Number of Lines
Core	6
L1 Cache	5-6
L2 and L3 Cache	6-7
Main Memory	5-6
Cache Geometry Definitions	5-15
Cache-Core & Cache-Cache Connection	4
Single Ring Bus Network Element	14

Table 2: Lines of code required per component

literature [3],[11], Taxim accepts plain text format for architecture generation and design parameter selection. Thus, these files can also be edited manually.

4.2 Simulation Preparation

In order to support a broad range of systems, Taxim provides automation bash scripts that can be executed on operating systems that support bash. Taxim provides input and output locations which Multi2Sim [18] and McPat [12] simulators read inputs and write output values. Automation scripts are initially in stencils that are found in the environment. They are processed after the designer manually completes them. As an example, in Figure 8, we present a stencil which relates the scripts to the Design Entry files that are shown as Benchmarks, Topologies and Design Parameters. In the Design Entry step, the designer enters the locations of the mentioned design entry files, thus Taxim populates the stencil scripts. This design also allows the

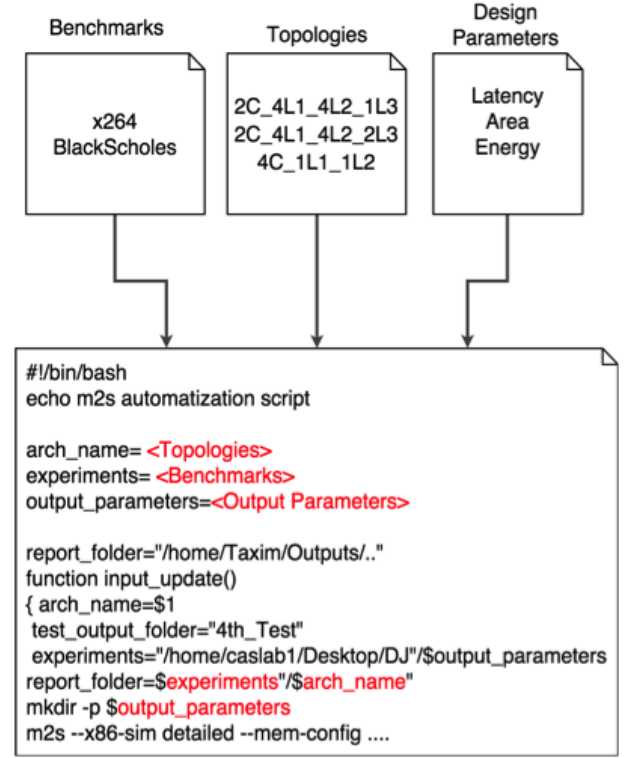


Figure 8: An example environment creation script used by Taxim to allow experimentation in succession

designer to conduct experiments in parallel by calling multiple instances in the terminal or adding related scripts in the same file.

4.3 Simulation and Intelligent Control

In this step, Taxim starts Multi2Sim [18] architectural simulation. Following the architectural simulation, power, area and timing simulations require McPAT [12] simulator. In order to customize McPAT [12], Multi2Sim simulation reports are extracted. McPAT [12] requires simulation output that represents all the components that are present in the architecture. These are datapath details of each core such as the number of instructions in the pipeline stages, cache and memory communication details such as cache miss/hit ratio, and interconnection network details such as sent and received packets etc. The output of McPAT [12] simulation are located in designer designated locations and the results are reported together with performance outputs. The designer does not need to do any additional work in order to extract McPAT [12] output. The output data extraction and reporting is explained in Section 4.5.

Taxim introduces two simple methods for early termination of simulation:

- While setting up the environment for multiple simulations, the designer might enter some parameters wrong. Hence, the simulations have to be aborted. Taxim prevents designers to simulate erroneous tests. When a very large number of experiments are configured with many benchmarks, input data sets and topologies, the designer can opt for first testing the

benchmarks, their input and the topologies by running a functional simulation that is a lot faster than cycle accurate simulation. If there exists no errors during functional simulation, then the cycle accurate simulation starts.

- The designer can provide a design parameter and a satisfaction condition. If the condition is not met after a predetermined simulation time which is chosen by the designer, Taxim drops the simulation and proceeds with the next topology in the simulation queue. For example, user can put a condition on Latency metric. If the desired Latency is not met after a certain amount of simulation time, Taxim stops simulating the design under test.

4.4 Reporting and Storage

Taxim extracts necessary output parameters from the simulation output files. Taxim recognizes all design parameters that Multi2Sim [18] and McPat [12] provide. Currently, Taxim supports Comma Separated Values (csv) and plain text file outputs to store these design parameters. The output locations and designated design parameters are configured by the designer in Step 1 that is shown in Figure 7. The generated results can be used for exploration [13], augmented to DSE tools [22] or stored in databases.

5. EXPERIMENTS

In this section, we present how our automated tool improves the overall simulation process that can be used as a standalone or part of a DSE tool. We have used Taxim in exploring various symmetric multiprocessing systems [13]. During our experiments in [13], 500 design points have been considered for simulation. These simulations involved around 100 different topologies and 5 different test cases from PARSEC [2] and MiBench [9] benchmarks.

5.1 Topology Generation Metrics

The lines of simulation codes generated by the topology generation algorithm is proportional to the topology. As an example, a complex topology such as 20C_10DL1_4IL1_5DL2_2L3_BP requires much more lines of code (LOC) for connectivity than a simple 2C_2DL1_2IL1_2DL2_2IL2_1L3 architecture. Taxim applies Algorithm 1 in order to build Multi2Sim [18] topology configuration file of the 20-core architecture. Thus, all the components and connectivity structures are defined. In Table 2, we present the number of lines that is required to describe components and connections between them. Hence, increasing number of components and bypass topologies take much more lines of code to represent. According to Table 3, the sample 20-Core design takes 443 LOC. However, for 2-core architecture, it is only 106 LOC. Thus, the automation Taxim provides will be more visible and helpful for future non-regular complex topologies.

To further illustrate this point, we have created a basic and very complicated model of each topology method using Taxim. Creating a topology list took less than five minutes and all the topologies were created within 70 milliseconds. Manual construction would take five to thirty minutes for each basic topology and an indefinite time for very complicated topologies.

Comparing the LOC required by 13C_9L1_5L2_3L3 and 18C_9DL1_6IL1_3L2, which is only a 5 line difference, it

is understandable that the complexity between topologies is not entirely dependent on the amount of cores, but the amount of connections that have to be defined for further modules. Hence, 18C_6L1_3L2 is very easy to code, but 6C_4DL1_3IL1_2L2 is a bigger challenge due to varying input and output streams between caches. The same is true for complicated networks too, as 33C_23DL1_17IL1_12DL2_4L3_BP has a higher line count than 37C_28DL1_19IL1_13DL2_8IL2_5L3_BP. This is mainly due to their network files, since 33C_23DL1_17IL1_12DL2_4L3_BP 17 IL1 caches bypass second and third level and connect to memory, whereas in 37C_28DL1_19IL1_13DL2_8IL2_5L3_BP 8 IL2 bypass third level cache and connect to the memory. According to Table III, network switches that are used in bypass topologies consume the most lines of code, therefore 33C_23DL1_17IL1_12DL2_4L3_BP consumes more lines than 37C_28DL1_19IL1_13DL2_8IL2_5L3_BP.

Table 3 also illustrates that, managing even simple topologies requires diligence for many connections, and creating topologies with higher complexity manually is prone to errors. Hence, by automating the creation and validation, we can eliminate mistakes, optimize the topology creation process and interconnections of topologies themselves.

Intelligent Simulation Control: The intelligent simulation control capabilities of Taxim that are explained in Section 4.3, have saved considerable amount simulation time in our experiments. As an example, Vips [2] benchmark on 4-core bypass and hybrid architectures takes 15 hours on average to complete in our simulation test bed. Using early termination method, we have opted to simulate Vips benchmarks on topologies which yield instructions per cycle (IPC) value greater or equal to 1.0. After thirty minutes of initial simulation, topologies that do not meet the IPC criteria have been eliminated. Thus, 10 topologies out of 64 had IPC less than 1.0. Remaining 54 topologies have been simulated with cycle accurate simulation. This elimination has reduced the overall simulation time by 12.2

Another aspect of Taxim's intelligent control is to prevent wasting simulation time due to erroneous design entry by the designer. This is overcome by first running functional simulation on each topology before the cycle-accurate simulation [18]. The functional simulation emulates underlying architecture and takes significantly less amount of time compared to detailed simulation time. In our testbed, x264 detailed simulation on a 4-core architecture takes 12 hours on average to complete whereas the functional simulation takes 6 minutes. Assuming that 60 simulations will be carried out sequentially, and 10 simulations are faulty due to erroneous design entry by the designer. If pre-check functional simulation had not been opted, 120 hours of simulation time would be wasted. However, by opting pre-check, the designer can fix the faults after functional simulations. Thus, in this experimental scenario, intelligent pre-check simulation saved 114 hours overall simulation time.

6. CONCLUSION

In this paper, we present Taxim, a toolchain that automates architectural, power, area and timing simulation using Multi2Sim [18] and McPat [12] tools. Taxim provides significant speed up for creating processor architectures and output generation compared to manual methods. Decision tree based automatic topology generation allows generating complex multi core architectures from simple text represen-

Component	Basic Topology	LOC	Generation Time (ms)
Regular	2C_2L1_2L2_1L3	76	9.09
Semi - Hybrid	4C_4DL1_2IL1_1L2	83	8.9
Hybrid	3C_3DL1_3IL1_3DL2_1IL2_1L3	122	8.74
2nd Level Bypass	2C_2DL1_2IL1_1DL2_1L3_BP	130	9.24
3rd Level Bypass	4C_4DL1_2IL1_2DL2_2IL2_1L3_BP	182	9.28
Component	Complex Topology	LOC	Generation Time (ms)
Regular	13C_9L1_5L2_3L3	227	8.64
Semi - Hybrid	18C_9DL1_6IL1_3L2	232	8.65
Hybrid	17C_11DL1_8IL1_5DL2_3IL2_2L3	321	9.09
2nd Level Bypass	32C_23DL1_17IL1_12DL2_4L3_BP	1105	9.39
3rd Level Bypass	37C_28DL1_19IL1_13DL2_8IL2_5L3_BP	979	9.28

Table 3: Examples of for each topology, followed by the amount of lines and generation times

tations. Intelligent control mechanism embedded in Taxim prevents wasting simulation time. Taxim source code is available for use at <https://github.com/bouncaslab/TaXim/>. In our experiments with taxim, we have shown that complex multiprocessing architectures can get more than a thousand lines in Multi2Sim's representation format, which is very difficult and slow to generate manually. However, Taxim can generate complex architectures in milliseconds. Alongside automatic processor generation, intelligent control prevents designers to save valuable simulation time due to human errors.

7. ACKNOWLEDGMENTS

Funding from the Turkish Ministry of Development under the TAM Project, number 2007K120610 and Bogazici University Scientific Projects number 7060 was received.

8. REFERENCES

- [1] A. Bengueddach, B. Senouci, S. Niar, and B. Beldjilali. Energy consumption in reconfigurable mp soc architecture: two-level caches optimization oriented approach. In *Design and Test Symposium (IDT), 2013 8th International*, pages 1–6. IEEE, 2013.
- [2] C. Bienia, S. Kumar, J. P. Singh, and K. Li. The parsec benchmark suite: Characterization and architectural implications. In *Proceedings of the 17th international conference on Parallel architectures and compilation techniques*, pages 72–81. ACM, 2008.
- [3] N. Binkert, B. Beckmann, G. Black, S. K. Reinhardt, A. Saidi, A. Basu, J. Hestness, D. R. Hower, T. Krishna, S. Sardashti, et al. The gem5 simulator. *ACM SIGARCH Computer Architecture News*, 39(2):1–7, 2011.
- [4] H. Calborean, R. Jahr, T. Ungerer, and L. Vintan. Optimizing a superscalar system using multi-objective design space exploration. In *Proceedings of the 18th International Conference on Control Systems and Computer Science (CSCS), Bucharest, Romania*, volume 1, pages 339–346, 2011.
- [5] T. E. Carlson, W. Heirman, and L. Eeckhout. Sniper: exploring the level of abstraction for scalable and accurate parallel multi-core simulation. In *Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis*, page 52. ACM, 2011.
- [6] R. Chis, M. Vintan, and L. Vintan. Multi-objective dse algorithms' evaluations on processor optimization. In *Intelligent Computer Communication and Processing (ICCP), 2013 IEEE International Conference on*, pages 27–33. IEEE, 2013.
- [7] V. Desmet, S. Girbal, A. Ramirez, A. Vega, and O. Temam. Archexplorer for automatic design space exploration. *IEEE micro*, (5):5–15, 2010.
- [8] R. Devigo, L. Duenha, R. Azevedo, and R. Santos. Multiexplorer: A tool set for multicore system-on-chip design exploration. In *Application-specific Systems, Architectures and Processors (ASAP), 2015 IEEE 26th International Conference on*, pages 160–161. IEEE, 2015.
- [9] M. R. Guthaus, J. S. Ringenberg, D. Ernst, T. M. Austin, T. Mudge, and R. B. Brown. Mibench: A free, commercially representative embedded benchmark suite. In *Workload Characterization, 2001. WWC-4. 2001 IEEE International Workshop on*, pages 3–14. IEEE, 2001.
- [10] H. Javaid, A. Ignjatovic, and S. Parameswaran. Performance estimation of pipelined multiprocessor system-on-chips (mpsocs). *Parallel and Distributed Systems, IEEE Transactions on*, 25(8):2159–2168, 2014.
- [11] Z. J. Jia, A. D. Pimentel, M. Thompson, T. Bautista, and A. Núñez. Nasa: A generic infrastructure for system-level mp-soc design space exploration. In *Embedded Systems for Real-Time Multimedia (ESTIMedia), 2010 8th IEEE Workshop on*, pages 41–50. IEEE, 2010.
- [12] S. Li, J. H. Ahn, R. D. Strong, J. B. Brockman, D. M. Tullsen, and N. P. Jouppi. Mcpat: an integrated power, area, and timing modeling framework for multicore and manycore architectures. In *Microarchitecture, 2009. MICRO-42. 42nd Annual IEEE/ACM International Symposium on*, pages 469–480. IEEE, 2009.
- [13] G. A. Malazgirt, B. Kiyan, D. Candas, K. Erdayandi, and A. Yurdakul. Exploring embedded symmetric multiprocessing with various on-chip architectures. In *Embedded and Ubiquitous Computing (EUC), 2015 IEEE 13th International Conference on*, pages 1–8, 2015.
- [14] R. P. Mohanty, A. K. Turuk, and B. Sahoo. Performance evaluation of multi-core processors with

- varied interconnect networks. In *Advanced Computing, Networking and Security (ADCONS), 2013 2nd International Conference on*, pages 7–11. IEEE, 2013.
- [15] P. Rosenfeld, E. Cooper-Balis, and B. Jacob. Drsim2: A cycle accurate memory system simulator. *Computer Architecture Letters*, 10(1):16–19, 2011.
- [16] D. Tarjan, S. Thoziyoor, and N. P. Jouppi. Cacti 4.0. Technical report, Technical Report HPL-2006-86, HP Laboratories Palo Alto, 2006.
- [17] C. Thompson, M. Gould, and N. Topham. High speed cycle approximate simulation for cache-incoherent mpsoes. In *Embedded Computer Systems: Architectures, Modeling, and Simulation (SAMOS XIII), 2013 International Conference on*, pages 88–95. IEEE, 2013.
- [18] R. Ubal, J. Sahuquillo, S. Petit, and P. Lopez. Multi2sim: A simulation framework to evaluate multicore-multithreaded processors. In *Computer Architecture and High Performance Computing, 2007. SBAC-PAD 2007. 19th International Symposium on*, pages 62–68, 2007.
- [19] N. Ventroux, A. Guerre, T. Sassolas, L. Moutaoukil, G. Blanc, C. Bechara, and R. David. Sesam: An mpsoe simulation environment for dynamic application processing. In *Computer and Information Technology (CIT), 2010 IEEE 10th International Conference on*, pages 1880–1886. IEEE, 2010.
- [20] K. Yan and X. Fu. Energy-efficient cache design in emerging mobile platforms: the implications and optimizations. In *Proceedings of the 2015 Design, Automation & Test in Europe Conference & Exhibition*, pages 375–380. EDA Consortium, 2015.
- [21] G. Yessin, A.-H. Badawy, V. Narayana, D. Mayhew, T. El Ghazawi, et al. "cere": A cache recommendation engine: Efficient evolutionary cache hierarchy design space exploration. In *High Performance Computing and Communications, 2014 IEEE 6th Intl Symp on Cyberspace Safety and Security, 2014 IEEE 11th Intl Conf on Embedded Software and Syst (HPCC, CSS, ICSSS), 2014 IEEE Intl Conf on*, pages 566–573. IEEE, 2014.
- [22] V. Zaccaria, G. Palermo, F. Castro, C. Silvano, and G. Mariani. Multicube explorer: An open source framework for design space exploration of chip multi-processors. In *Architecture of Computing Systems (ARCS), 2010 23rd International Conference on*, pages 1–7. VDE, 2010.